FISH: A Financial Interactive System for Signal Highlighting

Ta-Wei Huang[1*], **Jia-Huei Ju**[1*], **Yu-Shiang Huang**[1], **Cheng-Wei Lin**[1],
Yi-Shyuan Chiang[2], and **Chuan-Ju Wang**[1]

[1]Research Center of Information Technology Innovation, Academia Sinica
[2]Department of Computer Science, University of Illinois, Urbana-Champaign
a0917589225@gmail.com, {jhjoo, yushuang, lcw.1997}@citi.sinica.edu.tw,
ysc6@illinois.edu, cjwang@citi.sinica.edu.tw

Abstract

In this system demonstration, we seek to streamline the process of reviewing financial statements and provide insightful information for practitioners. We develop FISH, an interactive system that extracts and highlights crucial textual signals from financial statements efficiently and precisely. To achieve our goal, we integrate pre-trained BERT representations and a fine-tuned BERT highlighting model with a newly-proposed two-stage classify-then-highlight pipeline. We also conduct the human evaluation, showing FISH can provide accurate financial signals. FISH overcomes the limitations of existing research and more importantly benefits both academics and practitioners in finance as they can leverage state-of-the-art contextualized language models with their newly gained insights. The system is available online at https://fish-web-fish.de.r.appspot.com/, and a short video for introduction is at https://youtu.be/ZbvZQ09i6aw.

1 Introduction

Financial statements document the business activities and financial performance of a company. For example, the 10-K fillings required by SEC[1] are regulatory documents required of all public companies and are typically composed of several sections each. Considerable time and human resources are needed to digest such long and complicated texts. Accordingly, efficient analysis of complex and condensed documents is critical for financial practitioners. In this work, we introduce FISH—a **F**inancial **I**nteractive System for **S**ignal **H**ighlighting—as an effective and efficient system to review financial reports.

One common scenario in practice is when a company's report has just been released: financial professionals such as financial analysts and accountants must skim through the report and quickly prepare a preliminary summary. However, some parts of the report are minor or even trivial due to the established structure formulated by regulators or similar writing patterns from the same accounting firms. That is, there are often only a few sentences in the report that need to be carefully reviewed and analyzed. Although many studies leverage textual data in financial reports to provide soft evidence to support financial analysis (Liu et al., 2018; Du et al., 2019; Juan et al., 2021), most existing systems or studies still lack interactivity and do not directly provide off-the-shelf signals; such solutions are thereby considered impractical for many real-world usage scenarios.

2016 (Target)	*Our most critical accounting policies relate to revenue recognition, inventory, pension and other post-retirement benefit costs, goodwill, other intangible assets and long-lived assets and income taxes.*
2015 (Reference)	*Our most critical accounting policies relate to revenue recognition, inventory, pension and other post-retirement benefit costs, goodwill, other intangible assets and long-lived assets and income taxes.*

Table 1: A pair of highly similar segments from ITEM 7 in the financial 10-K reports of the Estée Lauder Companies Inc. in 2016.

We first recognize two challenges in the literature concerning textual information in long and complicated financial reports: (1) Many parts of a financial report are minor or even trivial; (2) It is difficult to utilize coarse information in empirical applications. To address these challenges, we propose a multi-stage financial analysis pipeline composed of two modules: a *segment classifier* and a *segment highlighter*.

To tackle the first one, we leverage the year-to-year structure of the annually released financial statements of a company. For example, as shown in Table 1, we observe that the target and reference segments appear identical, showing that these texts provide rather minor information and can be ignored for further analyses. For this part, we inte-

[*]These authors contributed equally to this work.

[1]Securities and Exchange Commission

grate a segment classifier that calculates the similarities for text pairs between years (i.e., the target year and the year previous to it). Given such a year-to-year similarity comparison, all segments in the report for a target year are classified as one of three types: (1) new segments, (2) highly similar segments, and (3) revised segments.

For the second challenge, the segment highlighter module provides straightforward and fine-grained signals in segments identified as the third type—revised segments—which are considered those revised from segments in the reports of the reference year. Specifically, this module highlights words in such segments by predicting the word importance based on the semantic context of the financial report and the differences between the two segments in a year-to-year pair. To accomplish this, we adopt contextualized representations from the pre-trained language model (Devlin et al., 2019) and further fine-tune the proposed module with a supervised token classification task.

In this demonstration, we showcase FISH, an interactive system to help financial professionals effectively and efficiently skim through financial reports in a straightforward manner. FISH is technically supported by the proposed two-module pipeline. In particular, we use financial 10-K reports collected by Loughran and McDonald (2011) to demonstrate our idea. FISH better visualizes the segment classifications in a target-year report and provides fine-grained information highlighting the essential information for the revised segments for financial professionals to review and analyze carefully.

2 Background and Related Work

Traditionally, research on financial statements focuses on quantitative data such as stock prices or other financial metrics. Textual information such as operation calls and forward-looking statements in reports are rarely carefully considered in conventional finance literature. Pioneering studies in both finance and computer science literature first adopted statistical or machine learning methods to identify crucial information in text data in financial reports. For example, Loughran and McDonald (2011) compile a large amount of 10-K reports and construct a finance-specific sentiment lexicon. Moreover, Jegadeesh and Wu (2013); Tsai and Wang (2017) leverage the sentiment signals in textual data to investigate relations between quantitative and textual information. More recently, distributed representation techniques have been introduced to analyze financial reports (Tsai et al., 2016; Rekabsaz et al., 2017; Lin et al., 2021).

Recent advancements in natural language processing (NLP) techniques have made it possible to develop useful information systems that can analyze textual information in financial reports. For example, Liu et al. (2018) leverage variants of pre-trained word embedding models to identify financial risks and cues to support financial analysis. Du et al. (2019) integrate multiple representations of 10-K reports and further infuse financial sentiment aspects into word and sentence representations. HIVE (Juan et al., 2021) is an interactive system utilizing an attention mechanism to explore insights from financial reports. However, existing systems do not effectively address the two challenges mentioned earlier, nor do they utilize state-of-the-art and dominant deep contextualized language models such as BERT (Devlin et al., 2019) and its variants as their back-end engine.

3 Financial Data and Pre-processing

The Form 10-K Financial Statements. we used the Form 10-K filings collected from the Software Repository for Accounting and Finance,[2] where a Form 10-K is an annual report required by the U.S. SEC. Specifically, we used the 10-K filings ranging from 2011 to 2018, which comprise 63,336 filings from 12,960 public companies. To make the best use of the year-to-year information, we discarded companies for which the reports in some years are missing during the period; 3,849 companies ($3{,}849 \times 8 = 30{,}792$ reports in total) remained after this filtering. Note that in this study, we randomly sample 200 companies from the 3,849 companies with their annual reports for demonstration purpose.

Coherent Text Segments. Every 10-K annual report contains 15 schedules (e.g., Items 1, 1A, 1B, 2, 3, ..., 7, 7A, ..., 15).[3] Each item section in a report is typically composed of multiple paragraphs, to which we first applied the SpaCy API[4] to divide each paragraph into sentences as our smallest unit of text. Moreover, as coherent text segments have been claimed to be beneficial to some downstream tasks such as information re-

[2] https://sraf.nd.edu/sec-edgar-data/

[3] https://en.wikipedia.org/wiki/Form_10-K

[4] Sentencizer: https://spacy.io/api/sentencizer

	#Segments/Report	#Tokens/Segment
Sentence	1,743	36
Segment*	677	94
Paragraph	474	134

Table 2: Statistics of pre-processed reports of 200 sampled companies. The two columns report the average numbers of segments in a report and the average numbers of tokens in a sentence/segment*/paragraph, respectively, where * indicates the documents are segmented by the cross-segment BERT.

trieval and other NLP applications (Koshorek et al., 2018; Shtekh et al., 2018), we further integrated the cross-segment BERT (Lukasik et al., 2020), a state-of-the-art text segmentation model, for the final pre-processing. Note that a segment may contain more than one sentence and usually reflects the proper length for the BERT-based models; thus, in our system, we take "segments" to be a basic unit as the input of the two proposed modules for classification and fine-grained highlighting. Table 2 is an overview of pre-processed segments with different levels of granularity and other data statistics.

4 The Multi-stage Pipeline

The proposed multi-stage pipeline is composed of the segment classifier and the segment highlighter modules, both of which leverage contextualized text representations from BERT-based models (Devlin et al., 2019; Reimers and Gurevych, 2019). With this pipeline we seek to examine year-to-year signals from the 10-K filings of each given company. Specifically, our interactive system targets each company's 10-K filings for a certain year; the company's report from the previous year is regarded as the reference document (see Table 1).

4.1 Segment Classifier

To leverage the year-to-year structure of a company's 10-K filing, we first denote the set of text segments from a company's year-t report as $\mathcal{S}_t = \{s_t^1, s_t^2, \ldots, s_t^n\}$, where n denotes the number of segments in the reports. As $\mathcal{S}_t$ is a target-year report, $\mathcal{S}_{t-1}$ is treated as a reference document. Accordingly, we perform year-to-year text ranking by treating segments in the target report $s_t^i \in \mathcal{S}_t$ as our queries and segments in the reference report $s_{t-1}^j \in \mathcal{S}_{t-1}$ as our references. In particular, the segment classifier calculates the similarity of each pair of target-reference text segments as

$$\phi(s_t^i, s_{t-1}^j),$$

where ϕ is a proximity function. In this study, we adopt two approaches for similarity calculation to account for both syntactic and semantic similarities. First, we use ROUGE-2 to measure the syntactic similarity, capturing bi-gram patterns in financial sentences (Lin, 2004). For semantic similarity, we utilize the fine-tuned SentenceBERT model (Reimers and Gurevych, 2019) to calculate the cosine similarity of each target-reference pair.

In this demonstration, each target text segment $s_t^i \in \mathcal{S}_t$ is classified into different groups by adopting the following heuristic rules with pre-defined thresholds τ and ϵ:[5]

$$s_t^i \text{ type} = \begin{cases} 1 & \text{if } \max(\{\phi_{\text{Rouge}}(s_t^i, s_{t-1}^j) | s_{t-1}^j \in \mathcal{S}_{t-1}\}) < \tau \\ 2 & \text{if } \max(\{\phi_{\text{Rouge}}(s_t^i, s_{t-1}^j) | s_{t-1}^j \in \mathcal{S}_{t-1}\}) \geq \tau \\ & \text{AND } \phi_{\text{BERT}}(s_t^i, s_{t-1}^{j^*}) \geq \epsilon, \\ 3 & \text{if } \max(\{\phi_{\text{Rouge}}(s_t^i, s_{t-1}^j) | s_{t-1}^j \in \mathcal{S}_{t-1}\}) \geq \tau \\ & \text{AND } \phi_{\text{BERT}}(s_t^i, s_{t-1}^{j^*}) < \epsilon, \end{cases} \tag{1}$$

where $j^* = \text{argmax}_j(\{\phi_{\text{Rouge}}(s_t^i, s_{t-1}^j) | s_{t-1}^j \in \mathcal{S}_{t-1}\})$ denotes the segment in the reference document with the maximum $\phi_{\text{Rouge}}(s_t^i, \cdot)$ similarity. Thus, a segment in the target report can be categorized according to the above three types:

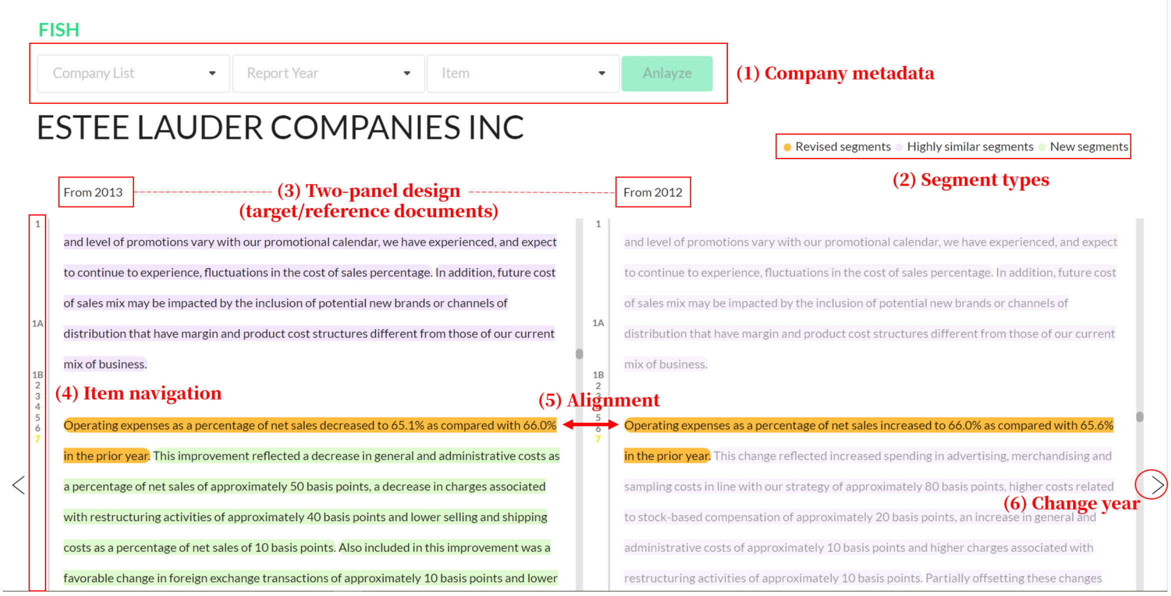
1. **New segments** are new text segments which are syntactically distant from all of their corresponding relevant reference text segments (as shown in Table 3).

2. **Highly similar segments** are text segments possessing syntactic structures and semantic meanings that closely resemble those of the reference segments (as shown in Table 1).

3. **Revised segments** include segments that are syntactically similar to the reference segments but differ semantically in meaning. In practice, as financial professionals shall pay greater attention to these segments, we here adopt further fine-grained highlighting for these segments in our second-stage module (as shown in Table 3).

Note that we here use a simple yet intuitive procedure to classify segments in target reports for

[5] We set threshold τ as 0.1, resulting in approximately 10% of new segments in a report; the threshold ϵ is set to 0.99, resulting in discarding approximately 50% of highly similar segments in a report.



From 2013:

and level of promotions vary with our promotional calendar, we have experienced, and expect to continue to experience, fluctuations in the cost of sales percentage. In addition, future cost of sales mix may be impacted by the inclusion of potential new brands or channels of distribution that have margin and product cost structures different from those of our current mix of business.

Operating expenses as a percentage of net sales decreased to 65.1% as compared with 66.0% in the prior year. This improvement reflected a decrease in general and administrative costs as a percentage of net sales of approximately 50 basis points, a decrease in charges associated with restructuring activities of approximately 40 basis points and lower selling and shipping costs as a percentage of net sales of 10 basis points. Also included in this improvement was a favorable change in foreign exchange transactions of approximately 10 basis points and lower

From 2012:

and level of promotions vary with our promotional calendar, we have experienced, and expect to continue to experience, fluctuations in the cost of sales percentage. In addition, future cost of sales mix may be impacted by the inclusion of potential new brands or channels of distribution that have margin and product cost structures different from those of our current mix of business.

Operating expenses as a percentage of net sales increased to 66.0% as compared with 65.6% in the prior year. This change reflected increased spending in advertising, merchandising and sampling costs in line with our strategy of approximately 80 basis points, higher costs related to stock-based compensation of approximately 20 basis points, an increase in general and administrative costs of approximately 10 basis points and higher charges associated with restructuring activities of approximately 10 basis points. Partially offsetting these changes

Figure 1: The main system interface of FISH. The left and right panels indicate the 10-K filings of a company in the target and reference years, respectively.

demonstration purposes; nevertheless, the classification can be much more complicated or involve professional adjustments of practitioners.

4.2 Segment Highlighter

The first-stage module narrows the considerations of what constitutes a (potentially) important segment. In the second module, we further focus on the third type of segments—the revised segments—and provide fine-grained information on these to enhance the readability of the documents. In particular, we seek to provide fine-grained (i.e., word-level) signals on such segments for practitioners, which in our demonstration of the interactive system is the basis for the highlighted words.

To build the highlighting model, we formulate the underlying word importance prediction problem as a token-level binary classification task by adding a classification linear layer on top of BERT. We further fine-tune the model using the e-SNLI dataset (Camburu et al., 2018), which was compiled for a natural language inference classification task that determines the entailment or contradiction relation for a given pair of sentences with human-annotated highlighted words. Fine-tuning takes around two hours on a V100 GPU, with less than 20GB of GPU memory usage.

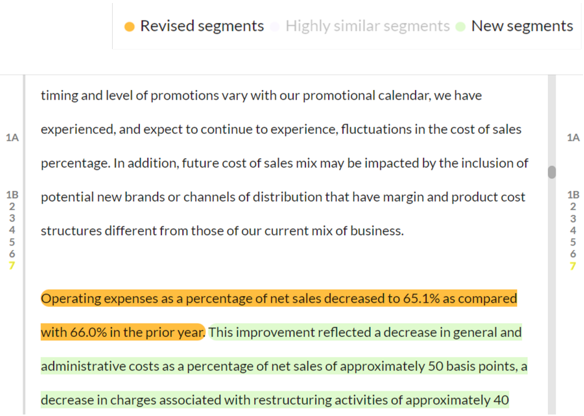
At the inference stage, for each syntactically similar but semantically dissimilar pair $(s^r_t, s^{j^*}_{t-1})$, where s^r_t is a revised segment, we construct the contextualized representation with BERT (Devlin et al., 2019) with the two special tokens (i.e., `[CLS]` and `[SEP]`) as:

$$h_{rj^*} = \mathrm{BERT}\big(\texttt{[CLS]}s^{j^*}_{t-1}\texttt{[SEP]}s^r_t\big). \qquad (2)$$

Recall that in Eq. (2), $s^{j^*}_{t-1}$ denotes the most syntactically similar segment in the reference year against s^r_t, but the cosine similarity between $s^{j^*}_{t-1}$ and s^r_t is rather low (see Eq. (1)). In this demonstration, we consider that word-level signals for such *revised* segments (syntactically similar but semantic dissimilar to the reference segments) can help users examining these segments easily and deeply. The importance probability of each word w in each revised segment s^r_t is then $P(w|s^r_t; s^{j^*}_{t-1}) = F(h_{rj^*}, w)$, where $F(\cdot,\cdot)$ denotes the fine-tuned model using the e-SNLI dataset; these probabilities are later used to indicate the word importance in our system using highlighting.

5 Demonstration

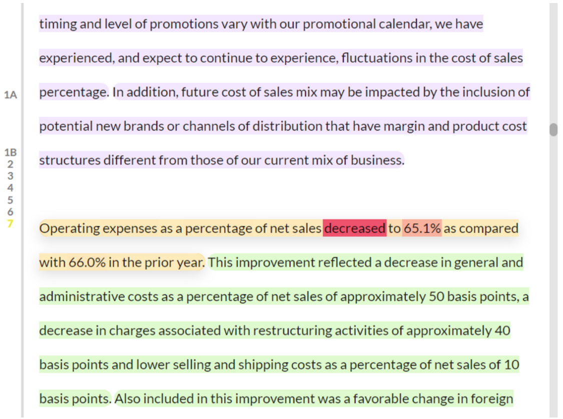
Figure 1 shows the main interactive interface of the proposed FISH, a web-based interactive analysis system for financial reports. For a better user experience, we lay out a concise system interface with user-friendly shortcuts to meet the needs of financial practitioners. In addition, we use both coarse-grained (segment-level) and fine-grained (word-level) signal highlighting features and interactive functions based on the proposed pipeline. FISH thereby facilitates more effective and efficient reviewing of financial reports.



timing and level of promotions vary with our promotional calendar, we have experienced, and expect to continue to experience, fluctuations in the cost of sales percentage. In addition, future cost of sales mix may be impacted by the inclusion of potential new brands or channels of distribution that have margin and product cost structures different from those of our current mix of business.

Operating expenses as a percentage of net sales decreased to 65.1% as compared with 66.0% in the prior year. This improvement reflected a decrease in general and administrative costs as a percentage of net sales of approximately 50 basis points, a decrease in charges associated with restructuring activities of approximately 40

Figure 2: Highlighting has been disabled for highly similar segments; only new or revised segments remain highlighted.



timing and level of promotions vary with our promotional calendar, we have experienced, and expect to continue to experience, fluctuations in the cost of sales percentage. In addition, future cost of sales mix may be impacted by the inclusion of potential new brands or channels of distribution that have margin and product cost structures different from those of our current mix of business.

Operating expenses as a percentage of net sales decreased to 65.1% as compared with 66.0% in the prior year. This improvement reflected a decrease in general and administrative costs as a percentage of net sales of approximately 50 basis points, a decrease in charges associated with restructuring activities of approximately 40 basis points and lower selling and shipping costs as a percentage of net sales of 10 basis points. Also included in this improvement was a favorable change in foreign

Figure 3: Segment- and word-level highlighting

5.1 User-friendly System Interface

The system includes the content of all available sections in the 10-K reports of the sampled 200 companies. As shown in (1) of Fig. 1, users can also scroll through drop-down menus at the top of the page to select the company and target year. Additionally, the ITEM navigation buttons on the left sidebar help users quickly locate the first line of the target ITEM, illustrated in (4) of the Fig. 1.

For year-to-year analysis, we adopt a two-panel interface (see (3) in Fig. 1) to make it easy for users to compare reports between consecutive years on the same screen. The left panel in the figure shows the contents of a financial report for the target year, and the right panel is regarded as the reference document and thus features a lower opacity. We also provide arrows on both sides of the screen by which to switch back and forth between target years, as shown in (6) of the Fig. 1.

5.2 Interactive Signal Highlighting

As described in Section 4, the proposed classify-then-highlight pipeline first classifies each segment in the target report as one of three types according to Eq. (1): new, highly similar, or revised. Each segment is highlighted in a color reflecting its type, as illustrated in the content panel of the figure. The three color indicator buttons on the right top of the page (shown in (2) of Fig. 1) allow users to enable/disable highlighting for each type of segment, as demonstrated in Fig. 2.

In addition to segment-level highlighting, we provide fine-grained information for revised segments. Recall that each *revised* segment s_t^r in the report of interest (displayed in the left panel) is further passed to the segment highlighter along with the most syntactical similar segment $s_{t-1}^{j^*}$ in the reference document (see the condition of the third type in Eq. (1) and the description in Section 4.2). The embedding of the $(s_t^r, s_{t-1}^{j^*})$ segment pair in Eq. (2) is then fed to the fine-tuned model $F(\cdot,\cdot)$ to estimate the word importance of each word in s_t^r. This importance is indicated with different color shades when users hover over the revised segments. As illustrated in Fig. 3, the words *decreased* and *65.1%* are darker than others, implying that these two words are judged to be more crucial than other words in the same segment.

Note that we additionally provide a segment alignment feature as shown in (5) of Fig. 1. This horizontally aligns the target segment s_t^i with the most syntactical similar segment $s_{t-1}^{j^*}$ from the reference report document for the highly similar and revised segments, where the right panel automatically redirects to the corresponding aligned segment in the reference document when the user clicks on such segments in the left panel.

6 Empirical Evaluation

In this section, we report real-world cases that FISH captured and evaluate FISH's highlighting results with human judgement.

Case studies We take the financial report of *Estée Lauder Companies Inc* in 2016 for example. Table 3 provides an example of *new segments* and *revised segments* along with their reference segments. Recall that the new segments capture content that is not syntactically similar to—or is less syntactically similar to—content from the previous year's document. As shown in the upper block in Table 3, the target segment (left) is identified

	Estee Lauder Cos. (2016) – target segment	*Estee Lauder Cos. (2015)* – reference segment
New	*On May 3, 2016, we announced a multi-year initiative (Leading Beauty Forward) to build on our strengths and better leverage our cost structure to free resources for investment to continue our growth momentum.*	*We also plan to continue to build upon and leverage our history of outstanding creativity, innovation and entrepreneurship in high quality products and services and engaging communications.*
Revised	*Based on this material weakness, the Company s management has concluded that, as of June 30, 2016, the Company s internal control over financial reporting was not effective...*	*Based on this assessment, the Company s management has concluded that, as of June 30, 2015, the Company s internal control over financial reporting was effective...*

Table 3: The cases of new and revised segment (left) with their corresponding reference segments s_{t-1}^{j*} (right).

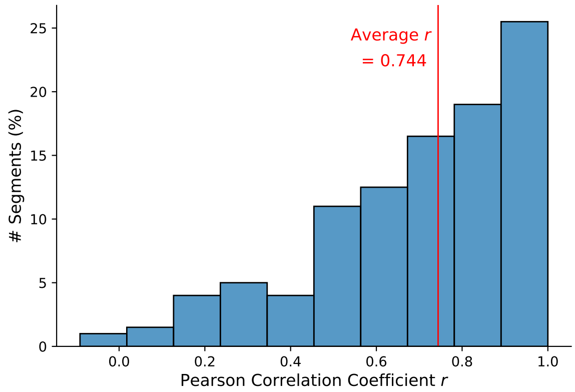


Figure 4: The histogram of the Pearson correlation coefficients r between human annotations and the outputs of our system. The red line is the average of all scores.

as a new segment by FISH as its corresponding most syntactically similar segment in the previous reference report is completely different. Indeed, with this example, we observe that the company is disclosing a new operational strategy in 2016, which is brand-new information compared to the previous year's report. As for the revised segments, which sometimes conceal important information such as changed income, expenses or management decisions such as new partnerships. In the lower block in Table 3, we observe what seems at first glance to be minor differences between reports in two consecutive years; however, the meanings behind these changes carry important financial signals (e.g. the highlighted words *weakness, was not effective*). With the highlighted words, we can then further attest the empirical effectiveness of these highlighted words.

Human Evaluation on Revised segments To verify the effectiveness of FISH's word-level highlighting on *revised* segments, we hire three assessors as potential users to select important words from the given segments. Specifically, the annotators should first identify the importance of each words-in-context of the revised segments and then label them as 1 or 0. As a result, for each sequence of words $[w_1, w_2, ...]$ in s_t^r, we calculate the Pearson correlation coefficient (denoted as Pearson's r, hereafter) between the human annotations[6] and the probabilities of word importance predicted by our system.

Our empirical evaluation data is composed of 200 *revised* segments randomly sampled from all revised segments classified by our system. As shown in Figure 4, most cases are with high values of Pearson's r, and only a few cases are with values lower than 0.5. Overall, FISH achieves a high average of 0.744 Pearson's r (the red vertical line in Figure 4).

7 Conclusion

We propose FISH, a financial statement signal-highlighting system integrated with a two-stage pipeline architecture, including a *segment classifier* and a *segment highlighter*. Both utilize BERT contextualized representations to strengthen the semantic comprehension of texts. Notably, our pipeline leverages the relationship of text segments between the target year and the previous year for automatic and interactive signal highlighting for financial professionals. The segment classifier first narrows the focus to new or revised segments instead of the entire report. As for the revised segments, we integrate a word-level highlighter to provide fine-grained financial signals via transfer learning on an external dataset. In addition, our human evaluation also suggests that FISH can provide effective highlighting results for empirical applications. To the best of our knowledge, FISH is the first interactive system not only made for practical financial applications but also leverages state-of-the-art contextualized language models, which shall greatly benefit both academics and finance practitioners to yield new insights.

[6] We take the average of three annotations as the final word importance (i.e., the ground truth) to avoid the personal subjective opinions.

References

Oana-Maria Camburu, Tim Rocktäschel, Thomas Lukasiewicz, and Phil Blunsom. 2018. e-snli: Natural language inference with natural language explanations. In *Proc. of NeurIPS*, pages 9539–9549.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional Transformers for language understanding. In *Proc. of NAACL*, pages 4171–4186.

Chi-Han Du, Yi-Shyuan Chiang, Kun-Che Tsai, Liang-Chih Liu, Ming-Feng Tsai, and Chuan-Ju Wang. 2019. Fridays: A financial risk information detecting and analyzing system. In *Proc. of AAAI*, pages 9853–9854.

Narasimhan Jegadeesh and Di Wu. 2013. Word power: A new approach for content analysis. *J. Financ. Econ.*, 110(3):712–729.

Yi-Ning Juan, Yi-Shyuan Chiang, Shang-Chuan Liu, Ming-Feng Tsai, and Chuan-Ju Wang. 2021. HIVE: Hierarchical information visualization for explainability. In *Proc. of IJCAI*, pages 4988–4991.

Omri Koshorek, Adir Cohen, Noam Mor, Michael Rotman, and Jonathan Berant. 2018. Text segmentation as a supervised learning task. In *Proc. of NAACL*, pages 469–473.

Chin-Yew Lin. 2004. Rouge: a package for automatic evaluation of summaries. In *Proc. of ACL, Text Summarization Branches Out*, pages 74–81.

Ting-Wei Lin, Ruei-Yao Sun, Hsuan-Ling Chang, Chuan-Ju Wang, and Ming-Feng Tsai. 2021. XRR: Explainable risk ranking for financial reports. In *Proc. of ECML-PKDD*, pages 253–268.

Yu-Wen Liu, Liang-Chih Liu, Chuan-Ju Wang, and Ming-Feng Tsai. 2018. RiskFinder: A sentence-level risk detector for financial reports. In *Proc. of NAACL*, pages 81–85.

Tim Loughran and Bill McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *J. Finance*, 66(1):35–65.

Michal Lukasik, Boris Dadachev, Kishore Papineni, and Gonçalo Simões. 2020. Text segmentation by cross segment attention. In *Proc. of EMNLP*, pages 4707–4716.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proc. of EMNLP-IJCNLP*, pages 3982–3992.

Navid Rekabsaz, Mihai Lupu, Artem Baklanov, Alexander Dür, Linda Andersson, and Allan Hanbury. 2017. Volatility prediction using financial disclosures sentiments with word embedding-based IR models. In *Procs. of ACL*, pages 1712–1721.

Gennady Shtekh, Polina Kazakova, Nikita Nikitinsky, and Nikolay Skachkov. 2018. Applying topic segmentation to document-level information retrieval. In *Proc. of CEE-SECR*. Article no. 6.

Ming-Feng Tsai and Chuan-Ju Wang. 2017. On the risk prediction and analysis of soft information in finance reports. *Eur. J. Oper. Res.*, 257(1):243–250.

Ming-Feng Tsai, Chuan-Ju Wang, and Po-Chuan Chien. 2016. Discovering finance keywords via continuous-space language models. *ACM Trans. Manage. Inf. Syst.*, 7(7):1–17.